DIAGNOSTIC PATHOLOGY

Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring

Anthony E Rizzardi[1], Arthur T Johnson[1], Rachel Isaksson Vogel[2], Stefan E Pambuccian[1], Jonathan Henriksen[1,3], Amy PN Skubitz[1,3], Gregory J Metzger[4] and Stephen C Schmechel[1,3*]

Abstract: Immunohistochemical (IHC) assays performed on formalin-fixed paraffin-embedded (FFPE) tissue sections traditionally have been semi-quantified by pathologist visual scoring of staining. IHC is useful for validating biomarkers discovered through genomics methods as large clinical repositories of FFPE specimens support the construction of tissue microarrays (TMAs) for high throughput studies. Due to the ubiquitous availability of IHC techniques in clinical laboratories, validated IHC biomarkers may be translated readily into clinical use. However, the method of pathologist semi-quantification is costly, inherently subjective, and produces ordinal rather than continuous variable data. Computer-aided analysis of digitized whole slide images may overcome these limitations. Using TMAs representing 215 ovarian serous carcinoma specimens stained for S100A1, we assessed the degree to which data obtained using computer-aided methods correlated with data obtained by pathologist visual scoring. To evaluate computer-aided image classification, IHC staining within pathologist annotated and software-classified areas of carcinoma were compared for each case. Two metrics for IHC staining were used: the percentage of carcinoma with S100A1 staining (%Pos), and the product of the staining intensity (optical density [OD] of staining) multiplied by the percentage of carcinoma with S100A1 staining (OD*%Pos). A comparison of the IHC staining data obtained from manual annotations and software-derived annotations showed strong agreement, indicating that software efficiently classifies carcinomatous areas within IHC slide images. Comparisons of IHC intensity data derived using pixel analysis software versus pathologist visual scoring demonstrated high Spearman correlations of 0.88 for %Pos ($p < 0.0001$) and 0.90 for OD*%Pos ($p < 0.0001$). This study demonstrated that computer-aided methods to classify image areas of interest (e.g., carcinomatous areas of tissue specimens) and quantify IHC staining intensity within those areas can produce highly similar data to visual evaluation by a pathologist.

Virtual slides: The virtual slide(s) for this article can be found here: http://www.diagnosticpathology.diagnomx.eu/vs/1649068103671302.

Keywords: Annotation, Color deconvolution, Digital pathology, Immunohistochemistry, Intensity, Quantification, Software

Despite the exceptional utility of genomics methods in the discovery phase of experimentation, these technologies require validation due to problems including misidentification of nucleic acid probes on gene expression microarrays [1,2], non-specificity of probes [3], and the essentially unavoidable false discovery rates associated with massive multiple hypothesis testing [4]. Appropriately powered studies to validate initial results of genomics studies often are lacking [5] or fail to confirm initial discovery-phase results [6], limiting clinical implementation of new disease biomarkers.

Immunohistochemistry (IHC) is an important technique for biomarker validation for several reasons. First, it allows direct visualization of biomarker expression in histologically relevant regions of the examined tissue. This is an important advantage over "grind and bind" assays in which tissue is solubilized for biochemical analysis, which may lead to false negative results if few biomarker-positive

* Correspondence: schme004@umn.edu
[1]Department of Laboratory Medicine and Pathology, University of Minnesota, 420 Delaware Street SE, MMC76, Minneapolis, MN 55455, USA
[3]BioNet, University of Minnesota, Minneapolis, MN, USA
Full list of author information is available at the end of the article

cells are present in a background of biomarker-negative tissue elements [7]. Second, clinical laboratories typically perform IHC on FFPE tissue sections processed by standard methods, making potentially available hundreds of millions of specimens for study [8]. Third, validated IHC assays may be implemented readily into clinical practice. For example, genomics methods were used to discover mRNA biomarkers capable of subclassifying diffuse large B cell lymphoma (DLBCL) into prognostically discrete subtypes [9]. Relevant subsets of these gene products were validated at the protein level using IHC on large numbers of DLBCL specimens [10,11], and validated IHC panels are now used clinically.

Traditionally, pathologists have visually scored IHC data. For example, in the calculation of an HSCORE, a summation of the percentage of area stained at each intensity level multiplied by the weighted intensity (e.g., 1, 2, or 3; where 0 is no staining, 1 is weak staining, 2 is moderate staining and 3 is strong staining) of staining is generated [12]. These analyses are frequently performed on specimens arrayed on stained TMA sections allowing representation of a sufficiently large number of specimens to for statistically rigorous testing [13,14]. Tissue specimens are adequately represented by tissue cores on very few slides [15,16] minimizing IHC cost and tissue usage, and facilitating intra-observer, inter-observer and inter-laboratory studies [10,17-20].

Pathologist visual scoring is fraught with problems due to subjectivity in interpretation. Automated IHC measurements promise to overcome these limitations. Whole-slide imaging systems are widely available to convert glass slides into diagnostic quality digital images [21]. Automated IHC measurements are precise in ranges of staining that appear weak to the eye [22] and produce continuous data [23]. Moreover, when automated IHC measurements are provided to a pathologist during visual scoring, computer-aided IHC analysis substantially improves both intra- and inter-observer agreement [20].

In this study, we used TMAs of ovarian serous carcinomas stained with an antibody directed against S100A1 to determine the ability of commercially available software algorithms (Genie Histology Pattern Recognition software suite including Genie Training v1 and Genie Classifier v1, and Color Deconvolution v9, Aperio Technologies, Vista, CA, USA) to replicate results obtained solely through visual inspection by a pathologist. Two specific comparisons were made in this study: a) the segmentation of the digitized tissue images into disease-relevant areas (those containing carcinoma) versus non-relevant areas (stroma and glass) and b) the quantification of stain intensity within areas of carcinoma. Specifically, first computer-derived IHC staining data obtained from both hand-annotated and Genie-classified areas of carcinoma were compared as a measure of agreement in tissue classification. Next, computer-derived IHC staining data from within Genie-classified areas of carcinoma were compared against pathologist visual scores.

Materials and methods
TMA Construction, IHC, and Pathologist visual scoring
Four TMA slides representing duplicate 0.6 mm core samples from 215 cases of ovarian serous carcinoma were provided by the Cheryl Brown Ovarian Cancer Outcomes Unit (Vancouver, Canada), stained with primary mouse anti-human S100A1 monoclonal antibody (clone DAK-S100A1/1; DakoCytomation, Glostrup, Denmark), and visualized with 3,3-diaminobenzidine (DAB) as previously described [24]. A total of 54, 54, 77 and 30 cases were represented by TMA 1, TMA 2, TMA 3, and TMA 4, respectively. Each TMA spot was examined by a pathologist (S.E.P.) who assigned a score of 0 (no staining), 1 (<10% of malignant cells staining), 2 (10%-50% of malignant cells staining), or 3 (>50% of malignant cells staining) within carcinomatous areas [24].

Slide digitization, Manual annotation, and Computer-aided image analysis
Digital images of IHC-stained TMA slides were obtained at 40x magnification (0.0625 μm^2 per raw image pixel) using a whole slide scanner (ScanScope CS, Aperio) fitted with a 40x/0.75 Plan Apo objective lens (Olympus, Center Valley, PA, USA). Images were saved in SVS format (Aperio), managed with server software (ImageServer, Aperio), and retrieved with a file management web interface (Spectrum, Aperio).

Under pathologist (S.C.S.) supervision, a technician (A.E.R.) hand-annotated tumor regions on whole slide images using Aperio's annotation software (ImageScope v10, Aperio). For automated image classification, image areas from TMA 1 were annotated that represented three user-defined Image Classes (carcinoma, stroma, and clear glass) and ranged in morphologic appearance and staining intensity of DAB and hematoxylin (counterstain). These image areas were used as input parameters for the histologic pattern recognition training software (Genie Training, Aperio) to produce a Genie Training Set. The effectiveness of the Genie Training Set was visualized on TMA 1 image test regions (TMA spots) using the image classifier algorithm (Genie Classifier, Aperio), which overlaid an image markup pseudocolored for each Image Class. Annotated image areas from TMA 1 were adjusted (adding or removing image areas) for each Image Class to improve the classifier accuracy. For example, if the Genie Classifier algorithm over-classified regions of stroma as carcinoma, additional stromal annotations were added to the Genie Training algorithm to better represent the stromal Image Class. This process of adjusting annotations, re-running the Genie Training algorithm, and visually inspecting pseudo-colored markup images output by Genie Classifier was

iteratively repeated until a Genie Training Set was developed to classify the TMA 1 slide optimally, as visually validated by a pathologist (S.C.S.). The optimized Genie Classifier was then run on TMAs 1-4.

IHC staining was evaluated within carcinomatous areas of each TMA spot that had been manually annotated, and a separate analysis was performed on areas from each TMA spot that had been classified as carcinoma by the Genie Classifier. As previously described [25,26], the Color Deconvolution algorithm (Aperio) was used to isolate individual stains for quantification: the red, green, and blue (RGB) OD color vectors were measured for each stain using default software settings and control slides stained separately with hematoxylin or DAB. The average RGB OD values (Hematoxylin: 0.682724, 0.642898, 0.347233; DAB: 0.486187, 0.588538, 0.645945) were entered into the Color Deconvolution software to define each stain component in the final analysis settings. Staining was quantified by two metrics: the percentage of carcinoma with S100A1 staining (%Pos), and the product of the staining intensity (OD) multiplied by the percentage of carcinoma with S100A1 staining (OD*%Pos). As previously described, the amount of staining present is linearly related to OD [26].
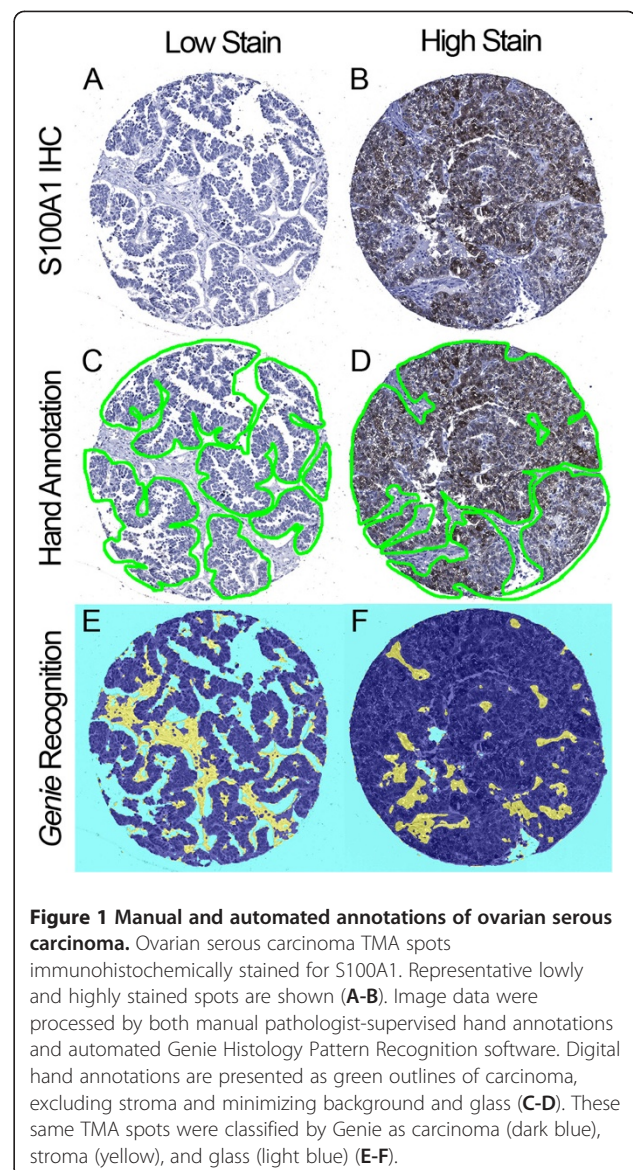
Statistical analysis

Duplicate spots were summarized as a single score for each case by randomly selecting one of the replicates. In order to compare pathologist hand and Genie automated annotations, which represent the same clinical measure on the same scale, Bland-Altman plots were used [27]. This scatterplot of the difference between methods, with reference lines at the mean difference and mean difference ± 2*-standard deviation of the differences, allows for an assessment of agreement rather than just a measure of correlation. Comparisons of both %Pos and OD*%Pos values by method were conducted. Spearman's correlation was calculated to compare pathologist visual scores versus %Pos and OD*%Pos values. Each comparison was made within each of the four TMAs. Additionally, we pooled all of the data to compare the %Pos and OD*%Pos values by pathologist score using Wilcoxon rank-sum tests.
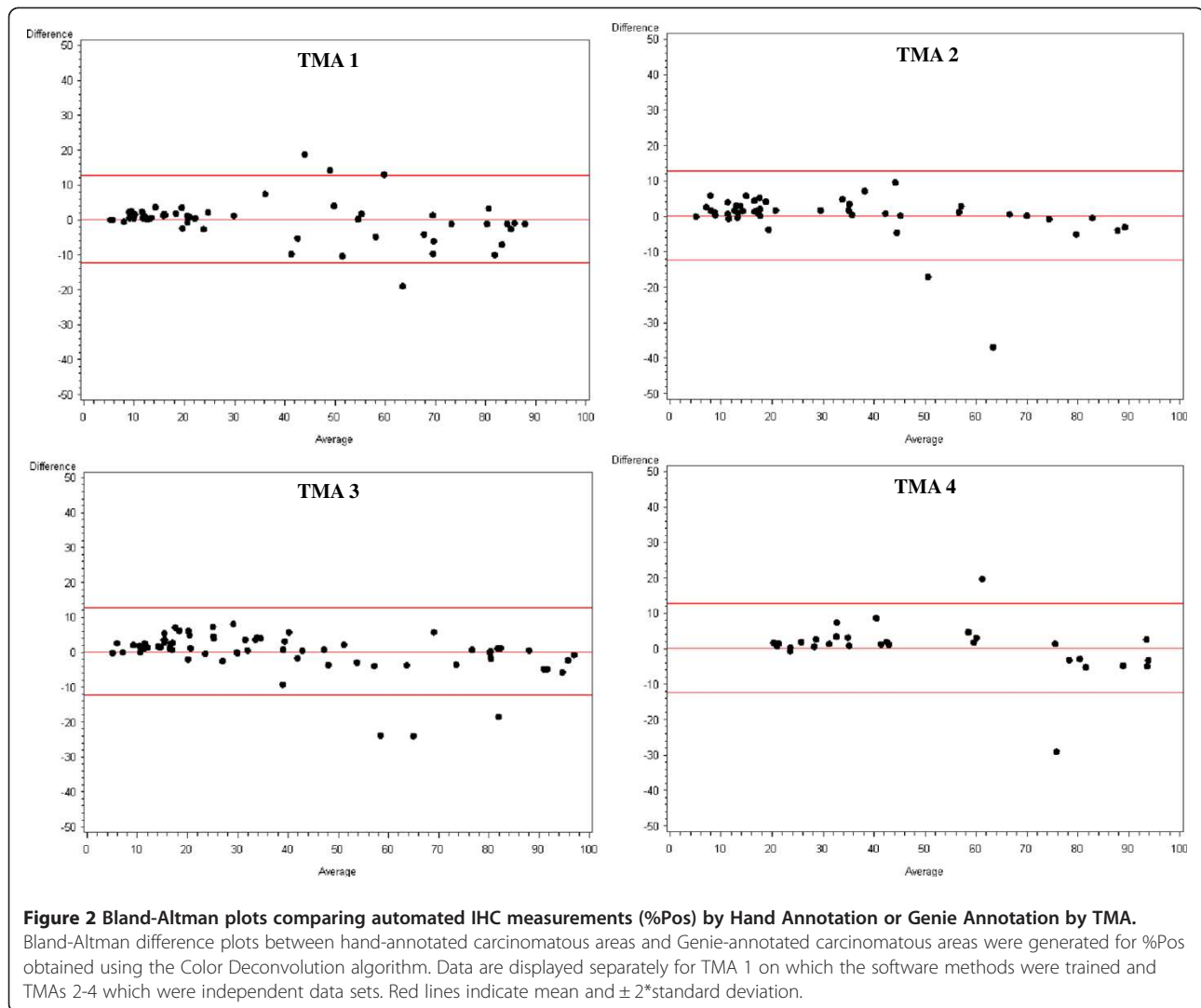
Results

Hand annotation versus Genie image classification of carcinoma

Representative TMA spots that had been stained for S100A1 by IHC were used for the analysis in this study are shown in Figure 1A,B. Examples of pathologist-directed, technician hand-annotation of areas of carcinoma, used in subsequent training and analysis, are shown in Figure 1C, D. The Genie Training Set algorithm was optimized and validated on TMA 1, a process that required one hour of pathologist time in addition to ten hours of technician

Figure 1 Manual and automated annotations of ovarian serous carcinoma. Ovarian serous carcinoma TMA spots immunohistochemically stained for S100A1. Representative lowly and highly stained spots are shown (**A-B**). Image data were processed by both manual pathologist-supervised hand annotations and automated Genie Histology Pattern Recognition software. Digital hand annotations are presented as green outlines of carcinoma, excluding stroma and minimizing background and glass (**C-D**). These same TMA spots were classified by Genie as carcinoma (dark blue), stroma (yellow), and glass (light blue) (**E-F**).

time. After optimization, the Genie Classifier algorithm was then run on all spots from TMAs 1-4 to classify areas of carcinoma, stroma and glass (Figure 1E,F). For both hand annotated and Genie classified carcinomatous areas, the Color Deconvolution algorithm was run to obtain %Pos and OD*%Pos data for DAB staining. The process of generating final data, which involved image quality control - for example to exclude damaged TMA spots from analysis - and organizing data output from Color Deconvolution, required an average of 3.5 hours per TMA, or 14 hours in total, of technician time.

There was strong agreement between data resulting from hand-annotation of carcinoma and data obtained after automated Genie classification of carcinoma (Figures 2 and 3). There was stronger agreement between the pathologist

Figure 2 Bland-Altman plots comparing automated IHC measurements (%Pos) by Hand Annotation or Genie Annotation by TMA.
Bland-Altman difference plots between hand-annotated carcinomatous areas and Genie-annotated carcinomatous areas were generated for %Pos
obtained using the Color Deconvolution algorithm. Data are displayed separately for TMA 1 on which the software methods were trained and
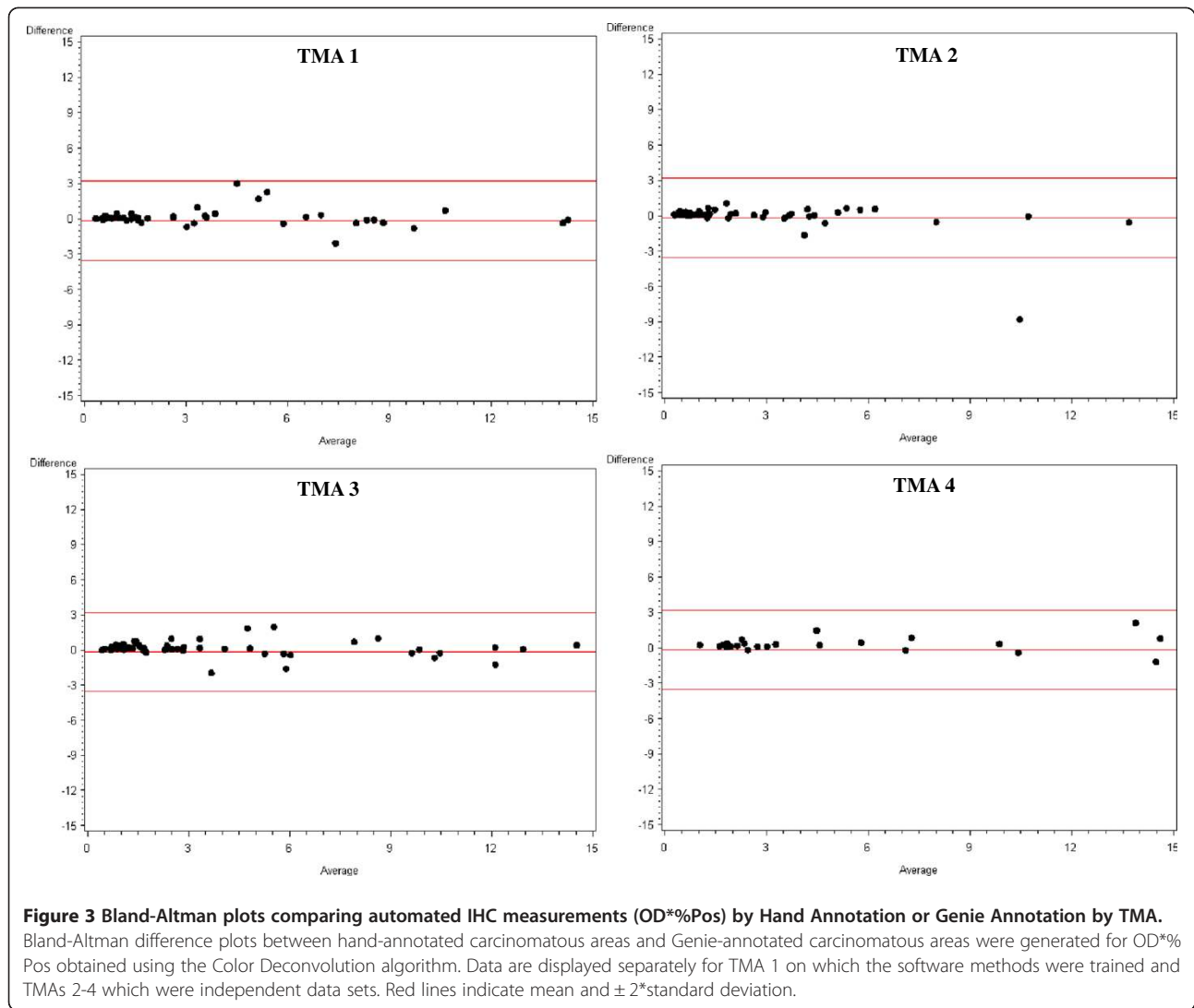TMAs 2-4 which were independent data sets. Red lines indicate mean and ± 2*standard deviation.

hand and automated Genie annotations for the OD*%Pos metric, evidenced by the lower variability in the mean difference in comparison with the %Pos metric.

Pathologist visual scoring in carcinoma versus Automated IHC measurement in Genie-classified carcinomatous areas
Using glass slides, a pathologist scored TMA spots for the percentage of positively stained carcinoma on a scale of 0-3+ as shown in representative spots covering the full scoring range in Figure 4A-D. For the 215 tumors in this study, scoring the TMA spots required 10 hours of pathologist time. In areas classified by Genie as carcinoma (Figure 4E-H), the Color Deconvolution algorithm individually analyzed DAB staining (deconvoluted by its RGB color components; Figure 4I-L) and %Pos and OD*%Pos data were obtained. As in Figure 1E,F, only the areas of carcinoma (pseudocolored as dark blue in Figure 1E,F and Figure 4E-H) were considered; areas

of stroma and glass (yellow and light blue, respectively, in Figure 1E-F and Figure 4E-H) did not contribute to the final IHC data. Data representative of OD*%Pos are illustrated as a heatmap in Figure 4M-P (gray = image areas not annotated by Genie as carcinoma and therefore not considered; blue = no staining, yellow = low intensities, orange = medium intensities, and red = high intensities in Genie-annotated carcinomatous areas considered). There was high correlation between pathologist visual scoring and %Pos data obtained using image analysis software for all TMAs, with Spearman correlations of 0.89, 0.78, 0.90 , and 0.90 for TMAs 1, 2, 3, and 4, respectively (all $p < 0.0001$; box plots of data shown in Figure 5). There was slightly higher correlation between pathologist visual scoring and OD*%Pos data, with Spearman correlations of 0.91, 0.81, 0.90, and 0.91, for TMAs 1, 2, 3, and 4, respectively (all $p < 0.0001$; box plots shown in Figure 6).
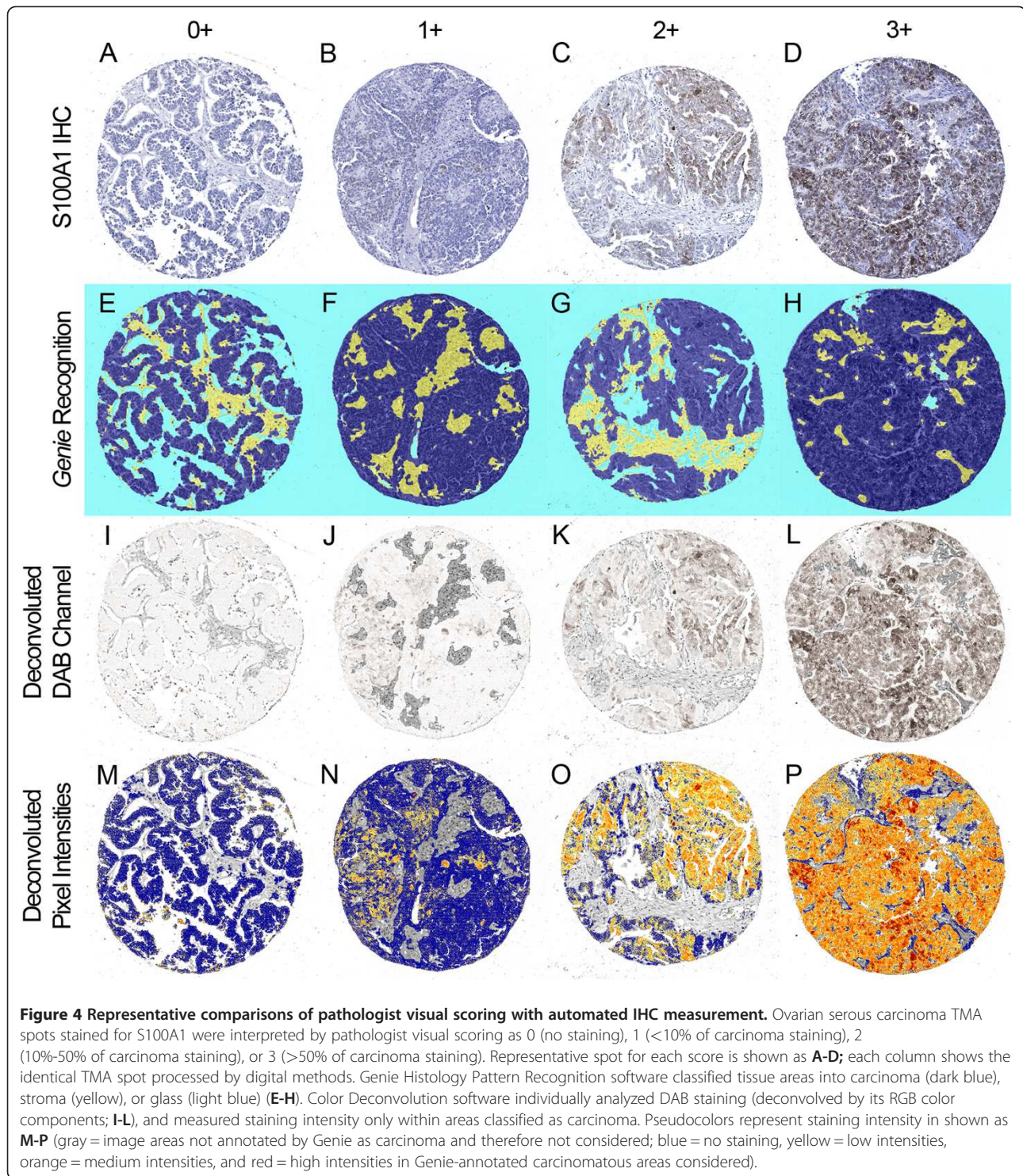
Figure 3 Bland-Altman plots comparing automated IHC measurements (OD*%Pos) by Hand Annotation or Genie Annotation by TMA.
Bland-Altman difference plots between hand-annotated carcinomatous areas and Genie-annotated carcinomatous areas were generated for OD*% Pos obtained using the Color Deconvolution algorithm. Data are displayed separately for TMA 1 on which the software methods were trained and TMAs 2-4 which were independent data sets. Red lines indicate mean and ± 2*standard deviation.

We next compared pathologist visual scoring with combined data (TMAs 1-4) from digital image analysis, revealing high correlation between pathologist visual scoring and %Pos (Spearman correlation 0.88, $p < 0.0001$) and OD*% Pos (Spearman correlation 0.90, $p < 0.0001$). There were significant differences in the median values for both metrics (%Pos and OD*%Pos) by pathologist score. Most notably, there were significant differences in computer-derived data corresponding to spots scored by the pathologists as "0" and "1" for both %Pos ($p < 0.0001$) and OD*% Pos ($p < 0.0001$).

Discussion

In this report we have demonstrated that commercially available software algorithms to classify disease-relevant tissue areas (Genie Histology Pattern Recognition) and quantify IHC staining within those areas (Color Deconvolution) effectively replicated IHC data produced by manual classification of image areas and pathologist visual scoring for S100A1 in ovarian serous carcinoma. Other software algorithms also provide data highly correlated with pathologist scores, e.g., human epidermal growth factor receptor 2 (HER2) [28-34], estrogen receptor [35-39] and progesterone receptor [37-39] in breast cancer, DNA mismatch repair proteins in esophageal cancer [40], and epidermal growth factor receptor signaling molecules in colon cancer [41], among other biomarkers.
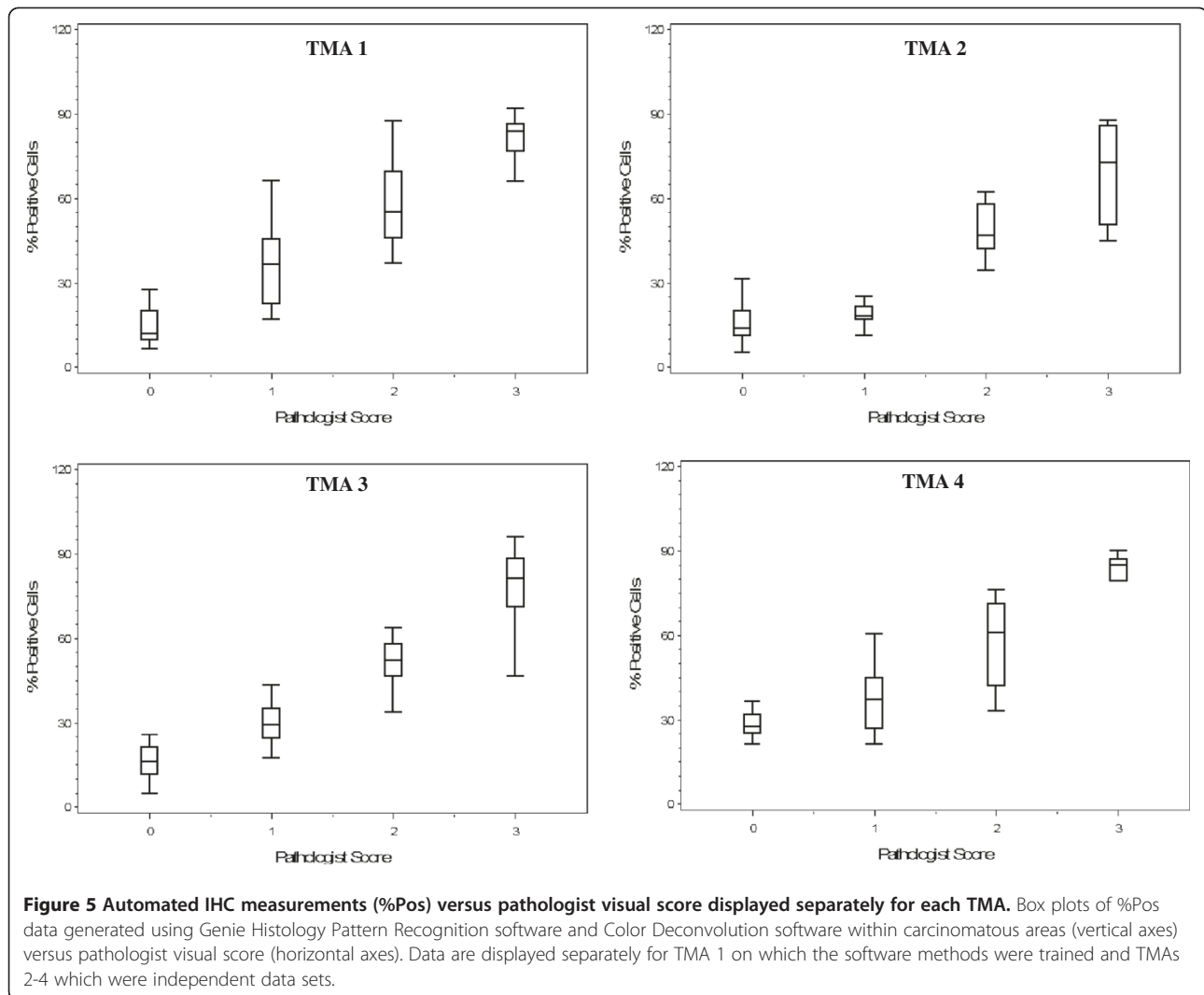
In this report, we provide important additional information regarding comparisons between digital data based solely on IHC-positive area (%Pos) and data combining area and staining intensity (OD*%Pos). The OD*%Pos metric provided better visual correlation between hand-annotated areas and Genie-annotated areas (Figure 4). Further, the OD*%Pos metric provided slightly higher correlation between digital IHC data and pathologist visual scoring. Of note, the study pathologist (S.E.P.) scored

Figure 4 Representative comparisons of pathologist visual scoring with automated IHC measurement. Ovarian serous carcinoma TMA spots stained for S100A1 were interpreted by pathologist visual scoring as 0 (no staining), 1 (<10% of carcinoma staining), 2 (10%-50% of carcinoma staining), or 3 (>50% of carcinoma staining). Representative spot for each score is shown as **A-D;** each column shows the identical TMA spot processed by digital methods. Genie Histology Pattern Recognition software classified tissue areas into carcinoma (dark blue), stroma (yellow), or glass (light blue) (**E-H**). Color Deconvolution software individually analyzed DAB staining (deconvolved by its RGB color components; **I-L**), and measured staining intensity only within areas classified as carcinoma. Pseudocolors represent staining intensity in shown as **M-P** (gray = image areas not annotated by Genie as carcinoma and therefore not considered; blue = no staining, yellow = low intensities, orange = medium intensities, and red = high intensities in Genie-annotated carcinomatous areas considered).

TMA spots for this study based on IHC-stained area as described in the Materials and Methods section, rather than by using a method such as HSCORE, which summated the percentage of area stained at each intensity level multiplied by the weighted intensity (e.g., 1, 2, or 3) [12]. Thus, it is unclear from our data why OD*%Pos performed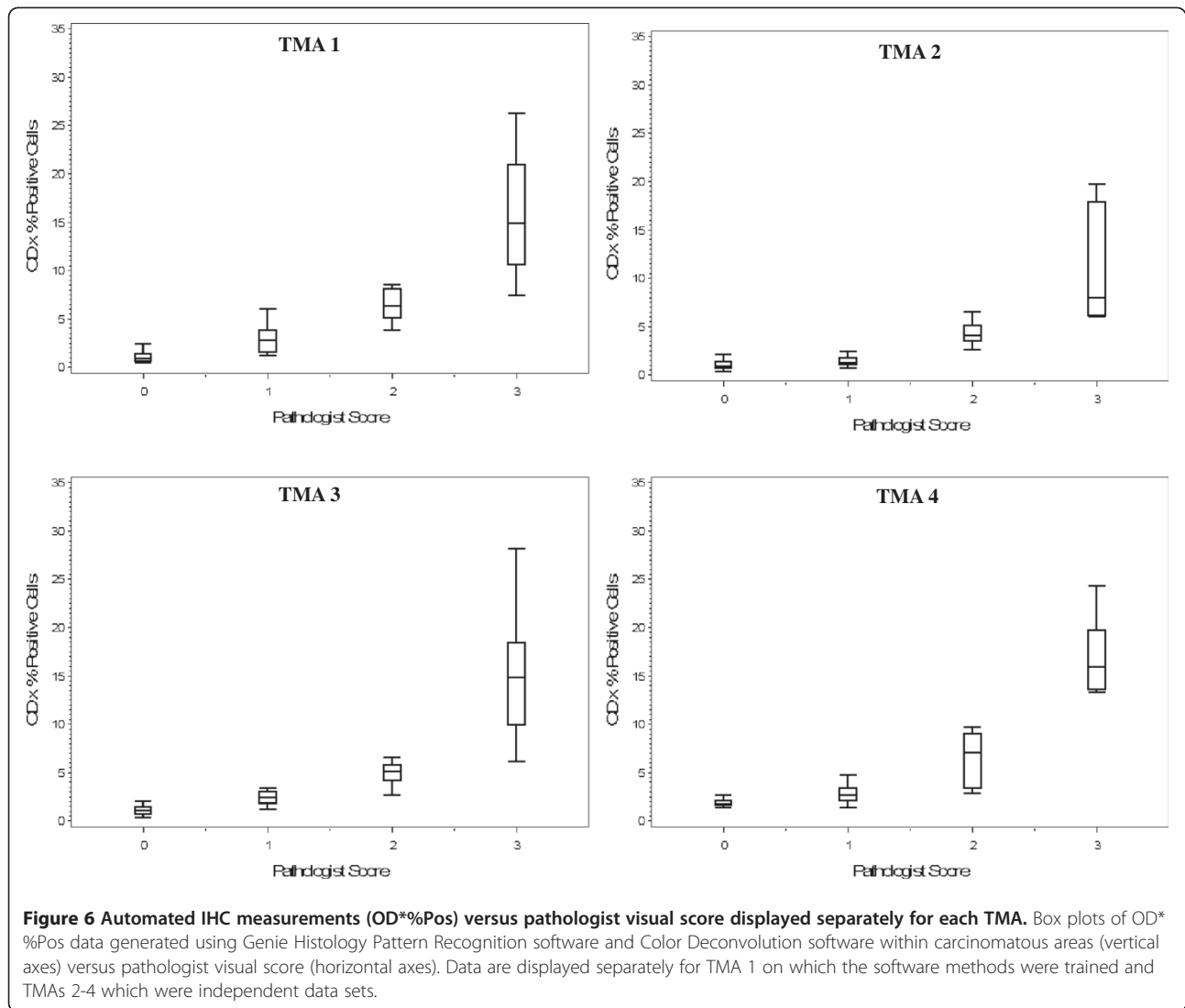 somewhat better than %Pos. We speculate that, since the human eye is more sensitive to higher intensity IHC staining [22], the estimation by eye of area IHC-stained likely inherently encompasses a component of staining intensity.

We additionally provide information regarding time conservation for pathologists using digital imaging methods for

Figure 5 Automated IHC measurements (%Pos) versus pathologist visual score displayed separately for each TMA. Box plots of %Pos data generated using Genie Histology Pattern Recognition software and Color Deconvolution software within carcinomatous areas (vertical axes) versus pathologist visual score (horizontal axes). Data are displayed separately for TMA 1 on which the software methods were trained and TMAs 2-4 which were independent data sets.

obtaining IHC data. While acknowledging that generating the automated IHC measurements within Genie-classified areas of carcinoma required 24 hours of technician time, 10-fold less pathologist time was required versus visual examination of each spot on TMAs 1-4. Greater efficiencies in the use of pathologists' time are needed as pathologists are experiencing increasing demands on their time due to higher clinical practice volumes, greater complexity of testing, and industry-wide shortages in available employees [42]. Although we did not measure pathologist time on a per-TMA spot basis in this study, a previous study indicates that per-spot time required for pathologist visual scoring of TMAs markedly increases as the number of spots to be analyzed increases [43]. Although limited data are available to assess pathologist fatigue on data quality, fatigue is postulated as a potential source of error in visual interpretation of IHC stained tissue sections [17]. To the contrary, automated analysis is objective and temporally linear regardless of the number of spots analyzed [43].

Although IHC biomarker studies widely use pathologist visual scoring, automated IHC measurement offers several additional advantages. First, pathologist visual scoring is fraught with data quality problems. The human eye is least accurate at detecting differences under conditions of weak staining at which IHC is most linearly related to target antigen concentration [22]. Consequently, regions of negative and high-positive intensities may be overcalled leading to artificially-produced bimodal score distributions [23]. While pathologist-derived data have good to excellent intra- and inter-observer reproducibility [18-20], estimation of percentages of areas stained has only poor to good reproducibility [19]. Digital methods may provide more reliable data. For example, automated HER2 IHC measurements are more comparable to consensus visual scores by multiple expert pathologists, and to HER2 gene amplification data, than are individual pathologists' subjective visual scores [44]. Since consensus scoring by experts is impractical in routine practice, automated IHC measurement may

Figure 6 Automated IHC measurements (OD*%Pos) versus pathologist visual score displayed separately for each TMA. Box plots of OD* %Pos data generated using Genie Histology Pattern Recognition software and Color Deconvolution software within carcinomatous areas (vertical axes) versus pathologist visual score (horizontal axes). Data are displayed separately for TMA 1 on which the software methods were trained and TMAs 2-4 which were independent data sets.

provide a means to improve IHC data quality. Intra- and inter-observer agreement is improved by providing pathologists with computer-aided IHC measurements during the visual scoring process [20,45]. Software algorithms such as Genie and Color Deconvolution may be "locked" such that all subsequent images are analyzed using the same parameters. Second, the automated methods demonstrated in this report also produced continuous variable data. Recent studies indicate that continuous variable data may allow identification of IHC cut-points of prognostic relevance that are either undetected [46] or are less statistically significant [23,34,47] by visual scoring. Third, digital methods support multigene expression studies at the protein level. Methods exist to multiplex IHC using immunofluorescence [48], destaining and restaining protocols [49], multiple chromagens [50,51], and combining data from

adjacent tissue sections [52,53]. Based on these and other studies, automated methods will likely become standard clinical practice.

Conclusions

This study demonstrated the effectiveness of optimized histology pattern recognition and automated IHC measurement algorithms to reproduce manual annotations and visual evaluation by a pathologist. This approach used TMAs in which tissue cores were obtained under the direction of a pathologist from areas containing exclusively tumor. A limited number of tissue cores adequately represent protein expression in tumor specimens [15,16]. Nevertheless, methods of quality control are required in final data analysis to exclude tissue areas with artifacts such as tissue folds, and tissue regions not of interest such as

admixed benign tissue elements in the analysis of carcinoma. It is important to note that we have found, in data not shown, that each combination of tissue type and IHC stain requires separate Genie optimization.

Abbreviations
CI: Confidence interval; DAB 3: 3-diaminobenzidine; FFPE: Formalin-fixed paraffin-embedded; HER2: Human epidermal growth factor receptor 2; IHC: Immunohistochemistry; OD*%Pos: Product of the staining intensity multiplied by the percentage of carcinoma with immunohistochemical staining; %Pos percentage of carcinoma with immunohistochemical staining; TMA: Tissue microarray.

Competing interest
The authors declare no conflict of interest.

Author details
[1]Department of Laboratory Medicine and Pathology, University of Minnesota, 420 Delaware Street SE, MMC76, Minneapolis, MN 55455, USA. [2]Biostatistics and Bioinformatics Core, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA. [3]BioNet, University of Minnesota, Minneapolis, MN, USA. [4]Department of Radiology, University of Minnesota, Minneapolis, MN, USA.

Authors' contributions
AER participated in study design, execution, analysis and interpretation of data, and drafting the manuscript. ATJ participated in study design and execution and analysis of data. RIV participated in study design, analysis and interpretation of data, and drafting the manuscript. SEP participated in execution of the study, interpretation of data, and reviewing the manuscript. JH assisted in execution of the study. APNS participated in execution of the study and reviewing the manuscript. GJM assisted in drafting the manuscript. SCS conceived of the study design, participated in data analysis and interpretation, and in drafting the manuscript. All authors read and approved the final manuscript.

References
1. Schmechel SC, LeVasseur RJ, Yang KH, Koehler KM, Kussick SJ, Sabath DE: Identification of genes whose expression patterns differ in benign lymphoid tissue and follicular, mantle cell, and small lymphocytic lymphoma. *Leukemia* 2004, 18:841–855.
2. Tu IP, Schaner M, Diehn M, Sikic BI, Brown PO, Botstein D, Fero MJ: A method for detecting and correcting feature misidentification on expression microarrays. *BMC Genomics* 2004, 5:64.
3. Kapur K, Jiang H, Xing Y, Wong WH: Cross-hybridization modeling on Affymetrix exon arrays. *Bioinformatics* 2008, 24:2887–2893.
4. Norris AW, Kahn CR: Analysis of gene expression in pathophysiological states: balancing false discovery and false negative rates. *Proc Natl Acad Sci U S A* 2006, 103:649–653.
5. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, Pfeiffer RM: Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst* 2005, 97:715–723.
6. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM: Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 2005, 97:1180–1184.
7. Cummings M, Iremonger J, Green CA, Shaaban AM, Speirs V: Gene expression of ERbeta isoforms in laser microdissected human breast cancers: implications for gene expression analyses. *Cell Oncol* 2009, 31:467–473.
8. Bouchie A: Coming soon: a global grid for cancer research. *Nat Biotechnol* 2004, 22:1071–1073.
9. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, *et al*: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, 403:503–511.
10. de Jong D, Xie W, Rosenwald A, Chhanabhai M, Gaulard P, Klapper W, Lee A, Sander B, Thorns C, Campo E, *et al*: Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications (a study from the Lunenburg Lymphoma Biomarker Consortium). *J Clin Pathol* 2009, 62:128–138.
11. Choi WW, Weisenburger DD, Greiner TC, Piris MA, Banham AH, Delabie J, Braziel RM, Geng H, Iqbal J, Lenz G, *et al*: A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res* 2009, 15:5494–5502.
12. McCarty KS Jr, Szabo E, Flowers JL, Cox EB, Leight GS, Miller L, Konrath J, Soper JT, Budwit DA, Creasman WT, *et al*: Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors. *Cancer Res* 1986, 46:4244s–4248s.
13. Camp RL, Neumeister V, Rimm DL: A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. *J Clin Oncol* 2008, 26:5630–5637.
14. Rimm DL, Camp RL, Charette LA, Costa J, Olsen DA, Reiss M: Tissue microarray: a new technology for amplification of tissue resources. *Cancer J* 2001, 7:24–31.
15. Camp RL, Charette LA, Rimm DL: Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000, 80:1943–1949.
16. Griffin MC, Robinson RA, Trask DK: Validation of tissue microarrays using p53 immunohistochemical studies of squamous cell carcinoma of the larynx. *Mod Pathol* 2003, 16:1181–1188.
17. Weaver DL, Krag DN, Manna EA, Ashikaga T, Harlow SP, Bauer KD: Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer. *Mod Pathol* 2003, 16:1159–1163.
18. Borlot VF, Biasoli I, Schaffel R, Azambuja D, Milito C, Luiz RR, Scheliga A, Spector N, Morais JC: Evaluation of intra- and interobserver agreement and its clinical significance for scoring bcl-2 immunohistochemical expression in diffuse large B-cell lymphoma. *Pathol Int* 2008, 58:596–600.
19. Jaraj SJ, Camparo P, Boyle H, Germain F, Nilsson B, Petersson F, Egevad L: Intra- and interobserver reproducibility of interpretation of immunohistochemical stains of prostate cancer. *Virchows Arch* 2009, 455:375–381.
20. Gavrielides MA, Gallas BD, Lenz P, Badano A, Hewitt SM: Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. *Arch Pathol Lab Med* 2011, 135:233–242.
21. Yagi Y, Gilbertson JR: A relationship between slide quality and image quality in whole slide imaging (WSI). *Diagn Pathol* 2008, 3(Suppl 1):S12.
22. Rimm DL: What brown cannot do for you. *Nat Biotechnol* 2006, 24:914–916.
23. Rimm DL, Giltnane JM, Moeder C, Harigopal M, Chung GG, Camp RL, Burtness B: Bimodal population or pathologist artifact? *J Clin Oncol* 2007, 25:2487–2488.
24. DeRycke MS, Andersen JD, Harrington KM, Pambuccian SE, Kalloger SE, Boylan KL, Argenta PA, Skubitz AP: S100A1 expression in ovarian and endometrial endometrioid carcinomas is a prognostic indicator of relapse-free survival. *Am J Clin Pathol* 2009, 132:846–856.
25. Ruifrok AC, Johnston DA: Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* 2001, 23:291–299.
26. Krajewska M, Smith LH, Rong J, Huang X, Hyer ML, Zeps N, Iacopetta B, Linke SP, Olson AH, Reed JC, Krajewski S: Image analysis algorithms for immunohistochemical assessment of cell death events and fibrosis in tissue sections. *J Histochem Cytochem* 2009, 57:649–663.
27. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986, 1:307–310.
28. Joshi AS, Sharangpani GM, Porter K, Keyhani S, Morrison C, Basu AS, Gholap GA, Gholap AS, Barsky SH: Semi-automated imaging system to quantitate Her-2/neu membrane receptor immunoreactivity in human breast cancer. *Cytometry A* 2007, 71:273–285.
29. Skaland I, Ovestad I, Janssen EA, Klos J, Kjellevold KH, Helliesen T, Baak JP: Comparing subjective and digital image analysis HER2/neu expression

scores with conventional and modified FISH scores in breast cancer. *J Clin Pathol* 2008, **61**:68–71.

30. Masmoudi H, Hewitt SM, Petrick N, Myers KJ, Gavrielides MA: **Automated quantitative assessment of HER-2/neu immunohistochemical expression in breast cancer.** *IEEE Trans Med Imaging* 2009, **28**:916–925.

31. Turashvili G, Leung S, Turbin D, Montgomery K, Gilks B, West R, Carrier M, Huntsman D, Aparicio S: **Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis.** *BMC Cancer* 2009, **9**:165.

32. Laurinaviciene A, Dasevicius D, Ostapenko V, Jarmalaite S, Lazutka J, Laurinavicius A: **Membrane connectivity estimated by digital image analysis of HER2 immunohistochemistry is concordant with visual scoring and fluorescence in situ hybridization results: algorithm evaluation on breast cancer tissue microarrays.** *Diagn Pathol* 2011, **6**:87.

33. Brugmann A, Eld M, Lelkaitis G, Nielsen S, Grunkin M, Hansen JD, Foged NT, Vyberg M: **Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains.** *Breast Cancer Res Treat* 2011, **132**:41–49.

34. Atkinson R, Mollerup J, Laenkholm AV, Verardo M, Hawes D, Commins D, Engvad B, Correa A, Ehlers CC, Nielsen KV: **Effects of the change in cutoff values for human epidermal growth factor receptor 2 status by immunohistochemistry and fluorescence in situ hybridization: a study comparing conventional brightfield microscopy, image analysis-assisted microscopy, and interobserver variation.** *Arch Pathol Lab Med* 2011, **135**:1010–1016.

35. Turbin DA, Leung S, Cheang MC, Kennecke HA, Montgomery KD, McKinney S, Treaba DO, Boyd N, Goldstein LC, Badve S, *et al*: **Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases.** *Breast Cancer Res Treat* 2008, **110**:417–426.

36. Gokhale S, Rosen D, Sneige N, Diaz LK, Resetkova E, Sahin A, Liu J, Albarracin CT: **Assessment of two automated imaging systems in evaluating estrogen receptor status in breast carcinoma.** *Appl Immunohistochem Mol Morphol* 2007, **15**:451–455.

37. Faratian D, Kay C, Robson T, Campbell FM, Grant M, Rea D, Bartlett JM: **Automated image analysis for high-throughput quantitative detection of ER and PR expression levels in large-scale clinical studies: the TEAM Trial Experience.** *Histopathology* 2009, **55**:587–593.

38. Krecsak L, Micsik T, Kiszler G, Krenacs T, Szabo D, Jonas V, Csaszar G, Czuni L, Gurzo P, Ficsor L, Molnar B: **Technical note on the validation of a semi-automated image analysis software application for estrogen and progesterone receptor detection in breast cancer.** *Diagn Pathol* 2011, **6**:6.

39. Bolton KL, Garcia-Closas M, Pfeiffer RM, Duggan MA, Howat WJ, Hewitt SM, Yang XR, Cornelison R, Anzick SL, Meltzer P, *et al*: **Assessment of automated image analysis of breast cancer tissue microarrays for epidemiologic studies.** *Cancer Epidemiol Biomarkers Prev* 2010, **19**:992–999.

40. Alexander BM, Wang XZ, Niemierko A, Weaver DT, Mak RH, Roof KS, Fidias P, Wain J, Choi NC: **DNA Repair Biomarkers Predict Response to Neoadjuvant Chemoradiotherapy in Esophageal Cancer.** *Int J Radiat Oncol Biol Phys* 2011, in press.

41. Messersmith W, Oppenheimer D, Peralba J, Sebastiani V, Amador M, Jimeno A, Embuscado E, Hidalgo M, Iacobuzio-Donahue C: **Assessment of Epidermal Growth Factor Receptor (EGFR) signaling in paired colorectal cancer and normal colon tissue samples using computer-aided immunohistochemical analysis.** *Cancer Biol Ther* 2005, **4**:1381–1386.

42. Muirhead D, Aoun P, Powell M, Juncker F, Mollerup J: **Pathology economic model tool: a novel approach to workflow and budget cost analysis in an anatomic pathology laboratory.** *Arch Pathol Lab Med* 2010, **134**:1164–1169.

43. Ong CW, Kim LG, Kong HH, Low LY, Wang TT, Supriya S, Kathiresan M, Soong R, Salto-Tellez M: **Computer-assisted pathological immunohistochemistry scoring is more time-effective than conventional scoring, but provides no analytical advantage.** *Histopathology* 2010, **56**:523–529.

44. Skaland I, Ovestad I, Janssen EA, Klos J, Kjellevold KH, Helliesen T, Baak JP: **Digital image analysis improves the quality of subjective HER-2 expression scoring in breast cancer.** *Appl Immunohistochem Mol Morphol* 2008, **16**:185–190.

45. Bloom K, Harrington D: **Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy.** *Am J Clin Pathol* 2004, **121**:620–630.

46. Harigopal M, Barlow WE, Tedeschi G, Porter PL, Yeh IT, Haskell C, Livingston R, Hortobagyi GN, Sledge G, Shapiro C, *et al*: **Multiplexed assessment of the Southwest Oncology Group-directed Intergroup Breast Cancer Trial S9313 by AQUA shows that both high and low levels of HER2 are associated with poor outcome.** *Am J Pathol* 2010, **176**:1639–1647.

47. Camp RL, Dolled-Filhart M, King BL, Rimm DL: **Quantitative analysis of breast cancer tissue microarrays shows that both high and normal levels of HER2 expression are associated with poor outcome.** *Cancer Res* 2003, **63**:1445–1448.

48. Camp RL, Chung GG, Rimm DL: **Automated subcellular localization and quantification of protein expression in tissue microarrays.** *Nat Med* 2002, **8**:1323–1327.

49. Glass G, Papin JA, Mandell JW: **SIMPLE: a sequential immunoperoxidase labeling and erasing method.** *J Histochem Cytochem* 2009, **57**:899–905.

50. Olin MR, Andersen BM, Zellmer DM, Grogan PT, Popescu FE, Xiong Z, Forster CL, Seiler C, SantaCruz KS, Chen W, *et al*: **Superior efficacy of tumor cell vaccines grown in physiologic oxygen.** *Clin Cancer Res* 2010, **16**:4800–4808.

51. Dandrea MR, Reiser PA, Gumula NA, Hertzog BM, Andrade-Gordon P: **Application of triple immunohistochemistry to characterize amyloid plaque-associated inflammation in brains with Alzheimer's disease.** *Biotech Histochem* 2001, **76**:97–106.

52. Mucci LA, Pawitan Y, Demichelis F, Fall K, Stark JR, Adami HO, Andersson SO, Andren O, Eisenstein A, Holmberg L, *et al*: **Testing a multigene signature of prostate cancer death in the Swedish Watchful Waiting Cohort.** *Cancer Epidemiol Biomarkers Prev* 2008, **17**:1682–1688.

53. Metzger GJ, Dankbar SC, Henriksen J, Rizzardi AE, Rosener NK, Schmechel SC: **Development of multigene expression signature maps at the protein level from digitized immunohistochemistry slides.** *PLoS One* 2012, **7**:e33520.